Full-length Genomic RNA of Papaya Leaf-Distortion Mosaic Virus

BACKGROUND OF THE INVENTION

1.Field of the Invention

The present invention relates to the full-length genomic RNA of papaya leaf-distortion mosaic virus.

2.Description of the Related Art

A problem of a disease called papaya leaf-distortion mosaic disease has arisen in papaya plants in Subtropic areas, causing mosaic symptoms on leaves and ring spots on fruits. It has been shown that this disease is caused by infection with a papaya leaf-distortion mosaic virus (hereinafter referred to as "PLDMV"). PLDMV belonging to the genus Potyvirus of the family Potyviridae is in a string-like shape, and is approximately 800 nanometers in length. The virus is transmitted nonpersistently by aphids. Viral components include its genome consisting of RNA and periplastic proteins surrounding the RNA. The RNA genes contain nucleotide sequences encoding 10 types of proteins required for infection and replication: P1, HC-Pro, P3, 6K1, CI, 6K2, NIa-VPg, NIa-Pro, NIb and CP.

Of these 10 types of proteins encoded by PLDMV genes, only the CP region encoding a periplastic protein has been analyzed so far. No other regions have been analyzed and none of the nucleotide sequences of these regions have been reported.

SUMMARY OF THE INVENTION

The use of the nucleotide sequence of the full-length genomic RNA in addition to the CP region would be very useful in elucidating the functions and roles of PLDMV. Accordingly, the object of the present invention is to determine the nucleotide sequence of the full-length genomic RNA of PLDMV.

To solve the problems, we have determined the full-length nucleotide sequence by cDNA cloning for the entire gene region of PLDMV. Then, we have completed the invention by elucidating the gene structure of regions encoding various proteins from the nucleotide sequence.

Accordingly, the present invention relates to an RNA and a DNA, each of which comprises a nucleotide sequence as shown in SEQ ID NO: 1 (or a nucleotide sequence complementary to said nucleotide sequence), or a nucleotide sequence as shown in SEQ ID NO: 1 in which uracil is replaced by thymine (or a nucleotide sequence complementary to said nucleotide sequence), respectively.

The present invention further relates to a method for diagnosing infection with PLDMV in a plant, comprising determing whether the plant is infected with the virus by detecting an RNA fragment specific in the virus from the plant, wherein the RNA fragment corresponds to a part of a nucleotide sequence as shown in SEQ ID NO: 1.

The present invention further relates to a method for producing a PLDMV-resistant plant, comprising integrating a DNA fragment having a function to impart resistance against PLDMV into the plant, wherein the DNA fragment corresponds to a part of a nucleotide sequence as shown in SEQ ID NO: 1.

The present invention further relates to a method for producing a foreign protein in a plant comprising the steps of:

1) synthesizing cDNA from genomic RNA of PLDMV;

2) adding a nucleotide sequence encoding an amino acid sequence, which can be cleaved with protease derived from PLDMV, to the 5' terminus and the 3' terminus of a gene encoding said foreign protein to obtain a DNA fragment having the nucleotide sequence and a nucleotide sequence of the gene;

3) inserting the DNA fragment of 2) into the cDNA of 1);

4) preparing an RNA by allowing an RNA polymerase to act on the cDNA of 3); and

5) infecting a plant with the RNA of 4).

The present invention further relates to a protein selected from the group consisting of the following (a) to (c), and DNAs encoding them:

(a) a protein comprising an amino acid sequence as shown in SEQ ID NO: 4;

(b) a protein comprising an amino acid sequence as shown in SEQ ID NO: 4 having deletion, substitution, or addition of one or more amino acids, and having a protease activity to cleave peptide bonds between Gln-Ala, Gln-Ser, and Glu-Gly; and

(c) a protein derived from PLDMV encoded by a DNA which hybridizes to a DNA comprising a nucleotide sequence as shown in SEQ ID NO: 3 or a DNA complementary to said nucleotide sequence under stringent conditions, and having a protease activity to cleave peptide bonds between Gln-Ala, Gln-Ser, and Glu-Gly.

This specification includes part or all of the contents as disclosed in the specification and/or drawings of Japanese Patent Application No.2001-40523, which is a priority document of the present application.

DETAILED DESCRIPTION OF THE INVENTION

Hereinafter, the present invention will be described in detail.

(1) RNA and DNA

RNA and DNA of the present invention relate to the full-length genomic RNA of papaya leaf-distortion mosaic virus("PLDMV"), and each of them comprises a nucleotide sequence as shown in SEQ ID NO: 1 (or a nucleotide sequence complementary to said nucleotide sequences), or a nucleotide sequence as shown in SEQ ID NO: 1 in which uracil is replaced by thymine(or a nucleotide sequence complementary to said nucleotide sequences), respectively.

DNA of the invention can be obtained from a cDNA library that is synthesized from the viral RNA, or directly from the viral RNA by the RT-PCR method, using appropriate primers which is prepared based on the genetic information shown in SEQ ID NO: 1.

Alternatively, if the information is not used, the DNA of the invention can be obtained, for example, by the following method we have carried out, with modification as needed.

Firstly, viral particles are isolated and purified from leaves of PLDMV-infected *Cucumis metuliferus*, and then an RNA is extracted from the particles. Using the RNA as a template, cDNA is synthesized with oligo dT primers. The resulting cDNA is incorporated into a phagemide vector pT7Blue for transformation of *E.coli*, and thereby obtaining a cDNA library. Then, PCR is performed using the transformed *E.coli* as a template so as to examine the presence or absence of inserts, and select plasmids containing the cDNA which contains PLDMV gene. Next, the cDNA obtained as described above are cloned. Using the cloned plasmids, nucleotide sequences of the cDNA can be determined by the

method, such as dideoxy method. Of the obtained nucleotide sequences, a sequence closest to 5' end of the cDNA is used to prepare a primer. Repetition of the above-mentioned steps can yield a more upstream nucleotide sequence.

RNA of the present invention can be obtained by transcribing the DNA of this invention.

The DNA and RNA of the invention can be used for the diagnosis of infection with PLDMV, production of a PLDMV-resistant plant, and production of a foreign protein in a plant, as described below.

(2) Diagnosing infection with PLDMV in a plant

A method of the invention for diagnosing infection with PLDMV is a method which comprises determing whether the plant is infected with the virus by detecting an RNA fragment specific in the virus from the plant, wherein the RNA fragment corresponds to a part of a nucleotide sequence as shown in SEQ ID NO: 1.

"an RNA fragment corresponds to a part of the nucleotide sequence as shown in SEQ ID NO: 1" as used herein means:

① the RNA fragment comprises a nucleotide sequence which is identical to a part of a nucleotide sequence as shown in SEQ ID NO: 1;

② the RNA fragment comprises a nucleotide sequence which is complementary to a part of a nucleotide sequence as shown in SEQ ID NO: 1;

③ the RNA fragment is that of ① or ②, having deletion, substitution, or addition of one or more nucleotides, and having species-specificity sufficient to use it as an index in diagnosing infection with PLDMV.

An RNA fragment to be detected may correspond to any region of a nucleotide sequence as shown in SEQ ID NO: 1, the RNA fragment

corresponding to P1 protein-coding region with high species-specificity is preferred. The P1 protein-coding region corresponds to a part of the sequence of the nucleotides 136 - 1575 as shown in SEQ ID NO: 1. It is demonstrated by BLASTN homology search that the region corresponding to the nucleotides 1 - 150 and the region corresponding to the nucleotides 1200 - 1440 of the P1 protein-coding region have high species-specificity, as shown in Fig 1. Thus, the RNA fragments correspond to the nucleotides 1 - 150 of the P1 protein-coding region and the RNA fragments correspond to the nucleotides 1200 - 1440 of the P1 protein-coding region are highly preferred.

A method for detecting an RNA fragment includes, but is not limited to, hybridization method using a labeled DNA or RNA as a probe, and RT-PCR method.

(3) A method for producing a PLDMV-resistant plant

A method for producing a PLDMV-resistant plant of the invention comprises integrating a DNA fragment having a function to impart resistance against PLDMV into a plant, wherein the DNA fragment corresponds to a part of a nucleotide sequence as shown in SEQ ID NO: 1.

"DNA fragment corresponds to a part of a nucleotide sequence as shown in SEQ ID NO: 1" as used herein means:

① the DNA fragment comprises a nucleotide sequence which is identical to a part of a nucleotide sequence as shown in SEQ ID NO: 1 in which uracil is replaced by thymine;

② the DNA fragment comprises a nucleotide sequence which is complementary to a part of a nucleotide sequence as shown in SEQ ID NO: 1 in which uracil is replaced by thymine; and

③ the DNA fragment is that of ① or ②, having deletion, substitution, or addition of one or more nucleotides, and having a function to impart resistance against PLDMV to the plant.

Tennant et al. have reported that they have succeeded in imparting virus resistance to a plant by integrating a region encoding a periplastic protein of papaya ringspot virus type P into the plant (Tennant et al., Phytopathology 84: 1359-1366, 1994). Maiti et al. have reported that they were able to impart virus resistance to a plant by integrating a region encoding a HC-Pro protein of tobacco vein mottling virus into the plant (Maiti, I.B., Murphy, J.F., Shaw, J.G., Hunt, A., 1993, Proc. Narl. Acad. Sci. USA. 90: 6110-6114). Further, Audy et al have reported that they were able to impart virus resistance to a plant by integrating a region encoding an NIb protein of potato virus Y into the plant (Audy, P., Palukaitis, P., Slack, S.A., Zaitlin, M., 1994, Molecular Plant-Microbe Inerractions 7: 15-22). Therefore, a preferable DNA fragment to be integrated into a plant corresponds to a part or whole of regions, including a capsid protein (CP) coding region (nucleotides 9064 - 9945 as shown in SEQ ID NO: 1), a HC-Pro coding region (nucleotides 1576 - 2949), and/or a NIb coding region (nucleotides 7501 - 9063). Furthermore, the part of these regions, including the regions corresponding to the nucleotides 1 - 380 and the nucleotides 780 - 882 of capsid protein (CP), the regions corresponding to the nucleotides 27 - 140 and the nucleotides 1280-1374 of a HC-Pro coding region, and/or the regions corresponding to the nucleotides 1 - 81 and the nucleotides 1447 - 1563 of a NIb coding region have high species-specificity. The results of BLASTN homology search are shown in Figs. 2 - 4. Therefore, the DNA fragments correspond to these regions are more preferable.

A PLDMV resistant plant can be produced by integrating a DNA fragment corresponding to a part of a nucleotide sequence as shown in SEQ ID NO: 1 into a plant cell with appropriate promoter and terminator sequences, and allowing the plant cell to regenerate to a plant body. A preferable plant cell, to which the DNA fragment is introduced, is derived from a PLDMV-infectious plant, including papaya, cucumber, *Cucumis melo* var. conomon, and *Cucumis metuliferus*. Examples of a form of the plant cell include, but are not specifically limited to, cultured cells, protoplasts, callus, slices of a leaf, embryos. Examples of a promoter sequence used herein include a 35S promoter of cauliflower mosaic virus, and an alcohol dehydrogenase 1 gene promoter. Examples of a terminator sequence used herein include a NOS terminator, and an alcohol dehydrogenase 1 gene terminator. Introduction of the DNA into the plant cell can be performed by various methods known to the skilled in the art. Examples of such a method include methods which use *Agrobacterium tumefaciens*, *Agrobacterium rhizogenes* and the like, an electroporation method, a polyethylene glycol method, and a particle gun method. A method for regenerating a plant cell to a plant body may be determined depending on a type of the plant cell. For example, when a plant is papaya, a method by Fitch et al. (Fitch, M. M. M., Manshardt, R. M., Gonsalves, D., Slightom, J. L., Sanford, J. C., 1992, Biotechnology 10: 1466-1472) can be used to regenerate the plant cell to a plant body.

(4) Production of a foreign protein in a plant

A method of the invention for producing a foreign protein in a plant comprises the following steps of 1) to 5).

1) cDNA is synthesized from genomic RNA of PLDMV. An example of the genomic RNA of PLDMV is an RNA comprising a nucleotide sequence as shown in SEQ ID NO: 1. Alternatively, an RNA comprising a nucleotide sequence as shown in SEQ ID NO: 1, having deletion, substitution, or addition of one or more nucleotides, and having infectious ability as a virus, may be used. cDNA can be synthesized by reverse transcription using a genomic RNA as a template. Here, the full-length genomic RNA or a part of the genomic RNA may be used as a template.

2) A nucleotide sequence encoding an amino acid sequence which can be cleaved with a protease derived from PLDMV is added to the 5' terminus and the 3' terminus of a gene encoding a foreign protein to be produced. Thus, the resulting DNA fragment includes both the nucleotide sequence and the gene. The gene encoding the foreign protein is not specifically limited and may be any gene. Examples of the amino acid sequence which can be cleaved with a protease derived from PLDMV include Gln-Ala, Gln-Ser, Glu-Gly, and the like. These amino acid sequences can be cleaved with NIa-Protease (hereinafter referred to as "NIa-Pro") derived from PLDMV.

3) The DNA fragment of 2) is inserted into the cDNA of 1). The DNA fragment of 2) may be inserted into any position between P3 region and CP region of the cDNA of 1). The gene encoding the foreign protein can be inserted with, e.g., restriction enzymes.

4) RNA polymerase is allowed to act on the resulting cDNA of 3), and thereby synthesizing an RNA.

5) The RNA of 4) is allowed to infect a plant.

(5) A protein having a protease activity

The proteins of this invention are selected from the group

consisting of the following (a) to (c):

(a) a protein comprising an amino acid sequence as shown in SEQ ID NO: 4;

(b) a protein comprising an amino acid sequence as shown in SEQ ID NO: 4 having deletion, substitution, or addition of one or more amino acids, and having a protease activity to cleave peptide bonds between Gln-Ala, Gln-Ser, and Glu-Gly; and

(c) a protein derived from PLDMV encoded by a DNA which hybridizes to a DNA comprising a nucleotide sequence as shown in SEQ ID NO: 3 or a DNA complementary to said nucleotide sequence under stringent conditions, and having a protease activity to cleave peptide bonds between Gln-Ala, Gln-Ser, and Glu-Gly.

The protein of (a) is NIa-Pro (a fragment having a protease activity of NIa) which was obtained from PLDMV used in the following Example 1. The amino acid sequence of NIa-Pro is shown in SEQ ID NO: 4 and the nucleotide sequence coding for NIa-Pro is shown in SEQ ID NO: 3. The nucleotide sequence as shown in SEQ ID NO: 3 corresponds to the nucleotides 6772 - 7500 as shown in SEQ ID NO: 1.

The protein of (b) is a protein in which mutation is introduced without decreasing or losing a protease activity of the original protein. Examples of such mutation include, but are not limited to, naturally-occurring and artificial mutations. An example of a technique to cause an artificial mutation is, but is not limited to, site-specific mutagenesis (see, Nucleic Acids Res. 10, 6487-6500, 1982). The number of amino acids mutated is not limited, provided that it does not lose a protease activity of the protein to cleave peptide bonds between Gln-Ala, Gln-Ser and Glu-Gly. Generally, the number is within 30 amino acids, preferably within 20 amino acids, more preferably

within 10 amino acids, and most preferably within 5 amino acids. The site of the protein responsible for the protease activity is G-x-C-G (Shukla, D.D., Ward, C.W. and Brunt, A. A. (1994) The potyviridae. CAB international, West Sussex.) which corresponds to the amino acids 149 - 152 (G-H-C-G) of NIa of PLDMV. Therefore, the mutation to the region except for the active site will not cause a lost of the protease activity, provide that the mutation will not change the conformation of the protein.

The protein of (c) is a protease derived from PLDMV which can be obtained by using a hybridization of DNAs. "Stringent conditions" used for the protein of (c) means conditions under which only specific hybridization occurs and non-specific hybridization does not occur. Such conditions are generally "1xSSC, 0.1%SDS, 37℃", preferably "0.5xSSC, 0.1%SDS, 42℃", more preferably "0.2xSSC, 0.1%SDS, 65℃". A DNA obtained by such hybridization generally shows high homology with a DNA comprising a nucleotide sequence as shown in SEQ ID NO: 3. The term "high homology" used herein means 60% or more of homology, preferably 75% or more of homology, and more preferably 90% or more of homology.

The proteins of the invention (proteins of (a) to (c)) have a protease activity to cleave peptide bonds between Gln-Ala (between Q-A), Gln-Ser (between Q-S), and Glu-Gly (between E-G). This can be presumed from the following.

The polyproteins of Potyvirus include 10 types of proteins, such as P1, HC-Pro, P3, 6K1, CI, 6K2, NIa-VPg, NIa-Pro, NIb, and CP. Of these proteins, P1 and HC-Pro has self-cleavage activity, P3 and the other proteins can be cleaved with NIa-Pro. That is, NIa-Pro has a function to recognize and cleave peptide bonds between P3-6K1, 6K1-CI,

CI-6K2, 6K2 - NIa-VPg, NIa-VPg - NIa-Pro, NIa-Pro - NIb, and NIb-CP. Table 1 shows amino acid sequences at the N terminus and at the C terminus of each protein composing the polyprotein of Potyvirus. As shown in the table, for PLDMV, there are three types of comnbinations of N-terminus amino acid of one protein and C-terminus amino acid of another protein: Gln and Ala (Q and A), Gln and Ser (Q and S), as well as Glu and Gly (E and G). Therefore, NIa-Pro from PLDMV is thought to cleave the peptide bonds between Gln-Ala, Gln-Ser, and Glu-Gly.

Table 1 also shows amino acid sequences at the N terminus and the C terminus of each protein composing the polyprotein of Potyviruses other than PLDMV. The cleavage sites of NIa-Pro derived from each virus other than PLDMV, which are presumed from datas in this table, are thought to be quite different from those of NIa-Pro derived from PLDMV.

Table 1

Literature in which sequences are described and

Accession numbers of Gen Bank

Virus	P1	/Hcpro	/P3	/6K1	/CI	/6K2	/NIa-Vpg/NIa-pro/NIb	/CP
PLDMV *1	M	Y/S	G/G	Q/A	Q/S	Q/S	E/G	E/G	Q/S	Q/S	Y
PVY *1	M	F/S	G/G	Q/R	Q/S	Q/A	Q/G	E/A	Q/A	Q/A	M
PepMoV *1	M	Y/S	G/G	Q/R	Q/S	Q/S	Q/G	E/A	Q/A	Q/S	M
TVMV *1	M	F/S	G/G	Q/A	Q/S	Q/S	Q/G	E/S	Q/G	Q/S	V
TEV *1	M	Y/S	G/G	Q/A	Q/S	Q/S	Q/G	E/G	Q/G	Q/S	Q
SbMV *1	M	Y/S	G/G	Q/A	Q/S	Q/S	Q/G	E/S	Q/G	Q/S	Q
PRSV *1	M	Y/N	G/G	Q/A	Q/S	Q/S	Q/G	E/G	Q/S	Q/S	N
PSbMV *1	M	F/S	G/G	Q/A	Q/S	Q/S	E/G	E/A	Q/S	Q/A	M
TuMV *1	M	F/S	G/G	Q/A	Q/T	Q/S	E/A	E/S	Q/T	Q/A	L
JGMV *1	M	Y/S	G/G	E/R	E/G	E/N	E/G	E/G	E/S	Q/S	I
PPV *1	M	Y/S	G/G	Q/S	Q/S	Q/T	Q/G	E/S	Q/S	Q/A	V
JYMV-JI *2	M	Y/S	G/G	Q/A	Q/A	Q/S	E/A	E/S	Q/M	Q/S	V
JYMV-M *3	M	F/A	G/G	Q/A	Q/G	Q/S	E/A	E/S	Q/M	Q/S	V
SPFMV *4	M	Y/S	G/G	Q/G	Q/S	Q/T	Q/G	E/S	Q/T	Q/S	V
RMV *5	M	Y/S	G/G	Q/A	Q/S	Q/S	E/G	E/S	Q/S	E/A	L
PSV *6	M	Y/S	G/G	Q/A	Q/S	Q/G	Q/G	E/S	Q/S	Q/S	Q
PVA *7	M	L/S	S/A	Q/A	Q/A	Q/S	Q/S	E/S	Q/G	Q/A	V

*1:Shukla, D.D., Ward, C.W. and Brunt, A.A. (1994). The potyviridae. CAB international, West Sussex., *2:AB016500, *3:AB027007, *4:NC 001841, *5:NC 001814, *6:NC 001723, *7:NC 001649

BRIEF DESCRIPTION OF THE DRAWING

Figure 1 shows the result of BLASTN homology search for P1 protein-coding region corresponds to a part of the sequence of the nucleotides 136 - 1575 as shown in SEQ ID NO: 1.

Figure 2 shows the result of BLASTN homology search for a capsid protein (CP) coding region (nucleotides 9064 - 9945 as shown in SEQ ID NO: 1).

Figure 3 shows the result of BLASTN homology search for a HC-Pro coding region (nucleotides 1576 - 2949).

Figure 4 shows the result of BLASTN homology search for a NIb coding region (nucleotides 7501 - 9063).

PREFERRED EMBODIMENTS OF THE INVENTION

Hereinafter, the present invention will be described more specifically by use of the following examples. However, the technical scope of the invention is not limited to these examples.

Example 1

Determination of the nucleotide sequence of PLDMV periplastic protein gene

(1) Isolation and purification of a virus

450 ml of 0.5M citrate buffer containing 0.56g of sodium sulfite (this buffer had been prepared with 0.5 M citric acid to pH 7.0) was added to 140 g of *Cucumis metuliferus* inoculated with PLDMV, and then ground with a blender. The homogenate was squeezed through cotton

cloth. Then, carbon tetrachloride was added to the filtrate, allowing the carbon tetrachloride to be 6% of the whole filtrate. After vigorous mixing, the filtrate was centrifuged at 6,000g and 4℃ for 15 min, so that the supernatant was obtained. To 500 ml of the supernatant, 37.6g of polyethylene glycol 6000, 2.92g of sodium chloride, 10ml of Triton x100 were added. The mixture was stirred at 4℃ for 90 min, and then centrifuged at 6,000g and 4℃ for 15 min. To the pellet precipitated after centrifugation, 0.1M citrate buffer containing 0.01M sodium sulfite (this buffer had been prepared with 0.1M citric acid to pH 7.0 and hereinafter referred to as a CD buffer) was added for re-suspension. The mixture was centrifuged at 6,000g and 4℃ for 15 min, thereby obtaining the supernatant. Next, 30ml of the supernatant was superposed over a 40% sucrose solution (prepared with CD buffer), and then centrifuged at 125,000g for 90 min. Then the pellet was resuspended with 20ml of a CD buffer, followed by centrifugation at 6,000g and 4℃ for 15 min, thereby obtaining the supernatant. Subsequently, 10ml of the supernatant was layered on 2ml of a 40% sucrose solution (prepared with a CD buffer), followed by centrifugation at 125,000g for 90 min. The pellet was resuspended with 2.5ml of a CD buffer, centrifuged at 6,000g and 4℃ for 15 min, thereby obtaining the supernatant. Then, the supernatant was layered on a linear density gradient of a cesium sulfate centrifugation (10-41%, Hitachi RPS40T rotor was used at 175,000g and 6℃ for 15 hours). Thus the obtained white band of a virus fraction was collected, diluted with a CD buffer, and then centrifuged at 238,000g and 4℃ for 90min. The precipitated virus pellet was resuspended with 0.3ml of 0.01M citrate buffer (pH 7.0), thereby obtaining a purified sample of the virus.

(2) Preparation of PLDMV-RNA

RNA was extracted from the purified PLDMV above using a commercially available nucleic acid extraction kit, Sepagene (Sanko Junyaku Co., Ltd.). Extraction was performed according to the attached instructions.

(3) Construction and screening of a cDNA library

Since the viral RNA belonging to the genus Potyvirus has a poly A sequence at its 3'terminus, a double-stranded cDNA was synthesized using an oligo dT primer. A series of steps was taken with a commercially available cDNA synthesis kit (CLONTECH) according to the instructions attached to the kit. Adapter primers were linked to both ends of the synthesized cDNA. Next, PCR was performed using a downstream primer (NIb1) which is complementary to a known sequence of the NIb protein region of PLDMV, and using an upstream primer (AP1) of a sequence contained in the adapter primer. The amplified product was subjected to column purification, and then inserted to a cloning site of a phagemide vector pT7Blue (Novagen). Column purification was performed using SizeSep400 Spum Columns (Amersham Pharmacia Biotech) according to the attached instructions. The reaction product was transferred into E.coli strain JM109.

A small amount of plasmids were rapidly prepared from the PLDMV cDNA library obtained as described above, thereby obtaining a clone (NIb-99) having an approximately 2Kb insert. The nucleotide sequence of the cDNA library was determined by the dideoxy method and analyzed with DNASIS (Hitachi Soft Engineering, Ver. 7.0).

Based on the upstream sequences of the determined nucleotide sequence, complementary primers were constructed. By repetition of

the above described PCR, cloning, and sequencing, each clone (NIa-41, CI-64, 6K1-46, HC-23, and P1-40) was obtained from downstream to upstream. Further, PCR was performed using primers complementary to sequences upstream of CI-64, primers homologous to sequences upstream of HC-23, and using cDNA library as a template. Thus, a clone (P16K1-11) having an approximately 4kb insert was obtained. The upstream sequence of PLDMV genome was determined from these clones.

(4) Determination of the 5' terminal sequence

Cloning of the 5' terminal portion of PLDMV gene has been tried several times by the 5' RACE method as described above. However, no plasmid containing this sequence was obtained. Then, primer extension was performed using the clone (P1-40) obtained in (3) above as a template, suggesting that 14 bases from the 5' terminus of PLDMV were not decoded yet. To elucidate the above sequence, improvement in the RNA purification method and the cloning method were tried.

TE (10mM Tris-HCl pH 8.0, 1mM EDTA) 68 μl, 10 μl of 20xSSC (3M NaCl, 0.3M sodium citrate pH 7.0), 2 μl of 20%SDS, and 20 μl of proteinase K (10mg/ml) were added to 100 μl of the purified PLDMV, and the mixture was kept at 37℃ for 60 min. Next, 100 μl of 0.5% bentonite solution, and 200 μl of TE saturated phenol solution were added to the mixture. Then the mixture was shaken and centrifuged with an eppendolf small type centrifuge for 3 min, thereby obtaining the aqueous layer. After repeating the phenol extraction process as described above, 200 μl chloroform was added to the aqueous layer. The mixture was shaken, centrifuged with an eppendolf small type centrifuge for 3 min, thereby obtaining the aqueous layer. To the thus obtained aqueous layer, 25 μl of 3M sodium acetate solution (pH 5.2),

and 500 μl of ethanol were added. The mixture was kept at -80℃ for 30 min, centrifuged with an eppendolf small type centrifuge for 10 min, thereby obtaining RNA as a precipitate. Next, 1 ml of 80% ethanol was added to the precipitate, followed by centrifugation with an eppendolf small type centrifuge for 3 min. Then, ethanol was removed, and RNA was dissolved in 100 μl of TE. In order to further increase purity of the RNA extract, the following steps were taken. 100 μl of 4M lithium chloride was added to the RNA solution, and then kept on ice for 4 hours, followed by centrifugation with an eppendolf small centrifuge for 10 min. 400 μl of 80% ethanol was added to the RNA precipitate, centrifuged for 3 min with an eppendolf small type centrifuge. After ethanol was removed, the RNA was dissolved in 12.5 μl of distilled water. Subsequently, 10 μl of 3M sodium acetate solution (pH 5.2) and 250 μl of ethanol were added to the mixture, kept at -80℃ for 30 min, and then centrifuged for 10 min with an eppendolf small type centrifuge, thereby obtaining RNA as the precipitate. One ml of 80% ethanol was added to the RNA, centrifuged for 3 min with an eppendolf small type centrifuge. After removal of ethanol, the RNA was dissolved in 10 μl of distilled water.

The cloning method was improved as follows. 1 μl of the complementary primer (P1-4) 100pM solution that had been prepared based on the sequence of the upstream portion of the clone (HC-23), 2 μl of the purified PLDMV-RNA above, and 7 μl of distilled water were mixed and kept at 65℃ for 5 min. Next, 9.2 μl of distilled water, 9.0 μl of 4xRT buffer (CLONTECH), 1.6 μl of 40U/μl RNase Inhibitor (CLONTECH), 3.7 μl of dNTPmix (10mM each), 0.5 μl of AMV Reverse Transcriptase (CLONTECH) were added to the solution, and then kept at 42℃ for 30 min. Thus ssDNA was synthesized. To this solution, 1 μ

1 of 0.5M EDTA (pH 8.0) was added and mixed, and then placed on ice. Subsequently, 2 μl of 6N NaOH was added to the mixture, and kept at 65℃ for 30 min. After RNA was degraded, 2 μl of 6N acetic acid was added to and mixed with the mixture, followed by addition of 16 μl of distilled water. DNA was purified from the solution using a QIA quick PCR Purification Kit (QIAGEN). Purification was performed according to the attached instructions.

The above ssDNA 2.5 μl was added with 2 μl of anchor primer (Zhi, 1996), 5 μl of 2xSingle-stranded Ligation Buffer (CLONTECH), 0.5 μl of 20U/μl T4 RNA Ligase (CLONTECH), and 0.5 μl of 50U/μl T4 RNA Ligase (TAKARA), and then allowed to stand at 22℃ overnight. Next, nested PCR was performed using this solution as a template, and a primer set (AP-B, P1-3) containing each sequence of the anchor primer and the complementary primer (P1-4) that had been used for reverse transcription reaction. Furthermore, nested PCR was performed using the reaction product as a template, and the more inward primer set (AP-C, P1-7). Then, cDNA was purified from the reaction product using a QIA quick PCR Purification Kit (QIAGEN), inserted into the cloning site of a phagemide vector pT7Blue (Novagen), thereby transferring into E.coli strain JM109. About 200 clones were selected from the cDNA library by colony PCR, thereby obtaining two clones (P1-7-6, P1-7-103) containing PLDMV 5' terminal sequences. Therefore, the 5' terminal sequence of PLDMV genome was decoded from these clones.

It was found that PLDMV genomic RNA comprised 10,155 bases, and had 6 bases of a poly A sequence at the 5' terminus followed by 129 bases of an untranslated region. There was an ORF starting from the initiation codon AUG at the 136th base from the 5' terminus and ending at the termination codon UAG at the 9943rd base. At the 3' terminus,

there was another untranslated region comprising 210 bases following a termination codon, and a poly A sequence existed following A at the 10,155th base, as well. It was also found that PLDMV genomic RNA might comprise 5 bases of a poly A sequence and 129 bases of an untranslated region at the 5' terminus. Furthermore, the untranslated region at the 3' terminus may comprise 209 bases, and a poly A sequence may exist following G.

A polyprotein encoded by ORF consisted of 3269 amino acids. With reference to Shukla et al.'s report (Shukla, D.D., Ward, C.W. and Brunt, A.A., 1994, The potyviridae, CAB international, West Sussex), the positions of various protein genes of PLDMV were specified. Therefore, it was shown that P1 consists of 480 amino acids (nucleotides 136 - 1575 as shown in SEQ ID NO: 1), HC-Pro of 458 amino acids (nucleotides 1576 - 2949 as shown in SEQ ID NO: 1), P3 of 348 amino acids (nucleotides 2950 - 3993 as shown in SEQ ID NO: 1), 6K1 of 52 amino acids (nucleotides 3994 - 4149 as shown in SEQ ID NO: 1), CI of 635 amino acids (nucleotides 4150 - 6054 as shown in SEQ ID NO: 1), 6K2 of 52 amino acids (nucleotides 6055 - 6210 as shown in SEQ ID NO: 1), NIa-VPg of 187 amino acids (nucleotides 6211 - 6771 as shown in SEQ ID NO: 1), NIa-pro of 243 amino acids (nucleotides 6772 - 7500 as shown in SEQ ID NO: 1), NIb of 521 amino acids (nucleotides 7501 - 9063 as shown in SEQ ID NO: 1), and CP of 293 amino acids (nucleotides 9064 - 9945 as shown in SEQ ID NO: 1), all of which are shown in SEQ ID NOs: 1 and 2.

All publications, patents and patent applications cited herein are incorporated by reference in their entirety.

Elucidation of various protein gene structures of PLDMV of this invention enables detection of PLDMV gene by the RT-PCR method using the primers which are constructed based on the gene sequence. For

example, there is a report that BYMV gene was detected from an infected plant by the RT-PCR method using primers that had been constructed based on the nucleotide sequence of bean yellow mosaic virus (BYMV) (Vunsh R, Rosner A, Stein A Ann Appl Biol 117: 561-569, 1990). Particularly, detection of P1 protein region with high species specificity allows highly accurate detection. For example, it has been reported that introduction of the periplastic protein gene of papaya ringspot virus type P (PRSV-P) into a papaya plant resulted in a virus-resistant plant (Tennant et al., Phytopathology 84; 1359-1366, 1994). That is, production of a PLDMV-resistant plant becomes possible by integrating the gene into the plant using genetic recombination techniques. Moreover, it has been reported that a foreign protein was produced in a plant body using an infectious clone of potato X virus or of tobacco mosaic virus as a vector (Ryabov, E.V. et al., Virology 242: 303-313, 1998). That is, insertion of a gene encoding a foreign protein into a PLDMV infectious clone allows use of the clone as an expression vector.